



Web Crawler Practice

Facebook Crawler



An-Hua Po¹, Dr. Chun-Hsiang Chan²

¹ Master Program in Intelligent Computing and Big Data, CYCU

² Department of Geography, National Taiwan Normal University



Outline

- 目標設定
- 環境及工具的準備
- FB爬蟲架構介紹
- 實戰



目標設定

目標設定：明確要做什麼

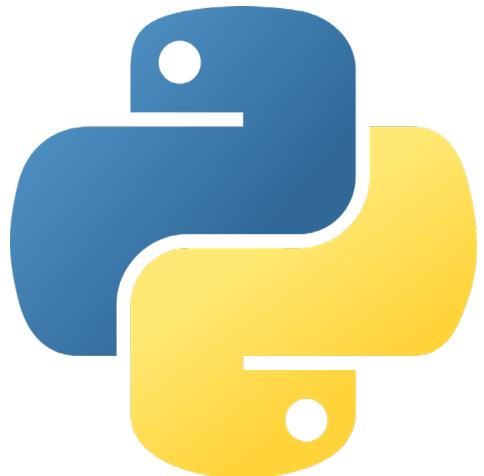


- (背景)你有一個曖昧對象，他/她是一個不折不扣的吃貨，你們已經曖昧一個月了
- 現在你有一個煩惱，**你不知道帶他/她去吃什麼好吃的了**，你知道好吃的地方都已經去過**(沒有好吃的就不和你曖昧，堅持三個月就和你在一起)**
- 於是你想到借助廣大網友的力量，去Facebook的美食社團找別人分享的好吃的
- 但是慢慢滑你覺得太麻煩，能不能利用Python爬蟲一次抓下來整理好呢？



環境及工具準備

環境、套件、工具



Python



Selenium



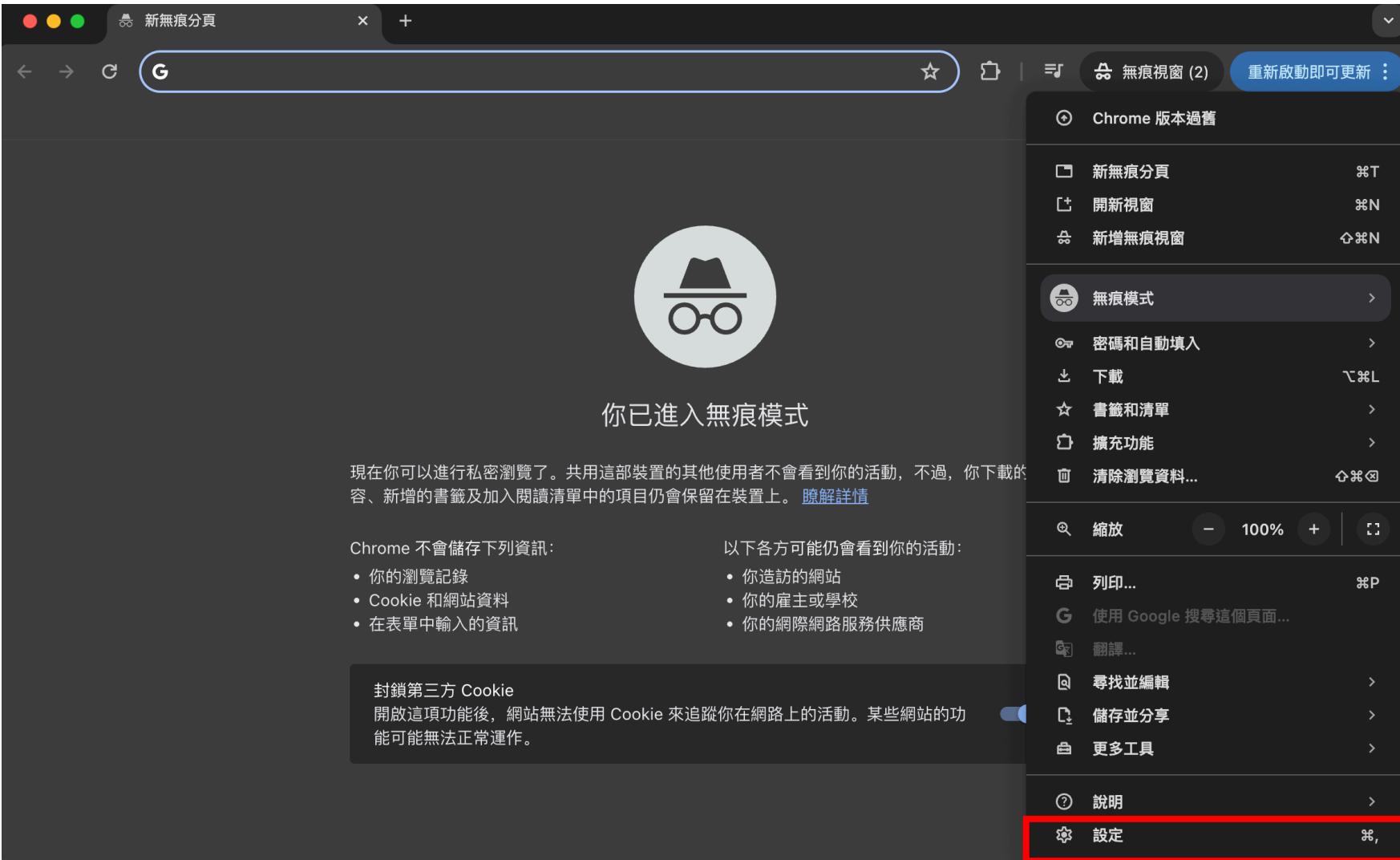
ChromeDriver



取得ChromeDriver

方法一：自己下載

確認Chrome版本



確認Chrome版本

The screenshot shows the Google Chrome settings interface. On the left, a sidebar lists various settings categories. At the bottom of this sidebar, the "關於 Chrome" (About Chrome) option is highlighted with a red box. The main content area is titled "關於 Chrome" and displays the following information:

Google Chrome
正在更新 Chrome
版本 121.0.6167.160 (正式版本) (arm64)

[前往 Chrome 說明頁面](#)

[回報問題](#)

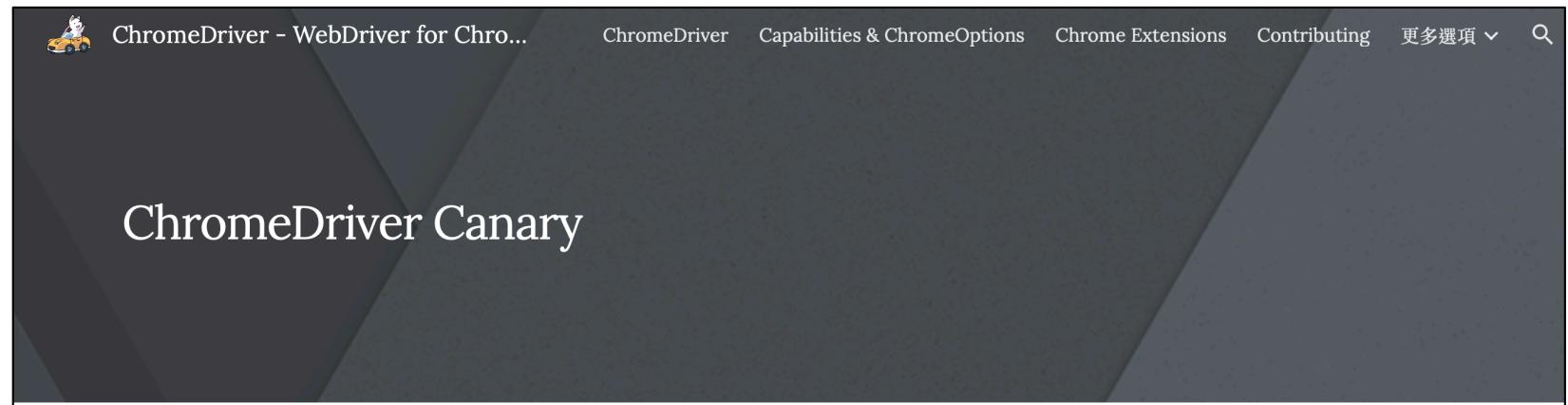
[隱私權政策](#)

Google Chrome
Copyright 2024 Google LLC. 保留所有權利。
Chrome 的開發仰賴 [Chromium](#) 開放原始碼計畫與其他 [開放原始碼軟體](#) 的協助才得以完成。
[服務條款](#)

打開ChromeDrive官網

■ ChromeDrive
下載網址：
<https://chromedriver.chromium.org/chromedriver-canary>

■ 注意作業系統
及位元數



ChromeDriver Canary has the latest new ChromeDriver features. New binaries are built and made available multiple times a day.

Be forewarned: it's designed for developers and early adopters, and can sometimes break down completely.

Starting with M115 the latest Chrome + ChromeDriver releases per release channel, including Canary, are available at [the Chrome for Testing availability dashboard](#). For automated version downloading one can use the convenient [JSON endpoints](#).

For older versions, you can find the canary builds at the following URLs:

- Linux (64-bit): https://commondatastorage.googleapis.com/chromium-browser-snapshots/index.html?prefix=Linux_x64/
- Mac OS X (64-bit): <https://commondatastorage.googleapis.com/chromium-browser-snapshots/index.html?prefix=Mac/>
- Mac ARM (64-bit): https://commondatastorage.googleapis.com/chromium-browser-snapshots/index.html?prefix=Mac_Arm/
- Windows (32-bit): <https://commondatastorage.googleapis.com/chromium-browser-snapshots/index.html?prefix=Win/>

it is generally a good idea to pick the latest build. In each of those folders, there is a file named LAST_CHANGE, that you can use to find the latest build number. Please go to one of the above URLs, depending on your system architecture, and type LAST_CHANGE in the filter box near the top. It may take the site some time to locate the file (up to a minute). When the LAST_CHANGE file link shows up, click on it to download the file, and find a build number inside it. Then type that build number into the filter box, and click on the folder link that shows up below.

Please note that the links above only provide ChromeDriver built on Feb 21, 2019 or later.

注意作業系統及位元數

確認系統與位元數後點擊對應網址

For older versions, you can find the canary builds at the following URLs:

- Linux (64-bit): https://commondatastorage.googleapis.com/chromium-browser-snapshots/index.html?prefix=Linux_x64/
- Mac OS X (64-bit): <https://commondatastorage.googleapis.com/chromium-browser-snapshots/index.html?prefix=Mac/>
- Mac ARM (64-bit): https://commondatastorage.googleapis.com/chromium-browser-snapshots/index.html?prefix=Mac_Arm/
- Windows (32-bit): <https://commondatastorage.googleapis.com/chromium-browser-snapshots/index.html?prefix=Win/>



下載ChromeDrive

Index of *chromium-browser-snapshots/Mac_Arm/*

Filter:				
Name	Last Modified	Size	Commit Position	Commit
1210557				
1210591				
1210611				
1210634				
1210651				
1210671				
1210683				
1210700				
1210718				
1210739				
1210752				
1210777				
1210807				

選擇與Chrome版本相近的即可

下載ChromeDrive

Index of chromium-browser-snapshots/Mac_Arm/1210634/

Filter:					
Name	Last Modified	Size	Commit Position	Commit	
Parent Directory			-	-	
REVISIONS	2023-10-17 05:04:11	661 bytes	-	-	
chrome-mac.zip	2023-10-17 05:04:08	129.23 MB	-	-	
chromedriver_mac64.zip	2023-10-17 05:04:13	8.45 MB	-	-	
content-shell.zip	2023-10-17 05:04:21	102.22 MB	-	-	
devtools-frontend.zip	2023-10-17 05:04:34	38.46 MB	-	-	
remoting-me2me-host-mac.zip	2023-10-17 05:04:36	53.49 MB	-	-	
updater.zip	2023-10-17 05:04:39	32.11 MB	-	-	

點擊會後自動下載

Python啟動ChromeDrive—測試

```
from selenium import webdriver
from selenium.webdriver.chrome.service import Service
driver_file = Service('../chromedriver_mac64/chromedriver') # 你下載ChromeDrive
的位置
driver = webdriver.Chrome(service=driver_file)
```



取得ChromeDriver

方法二 (推薦) : webdriver-manager套件

安裝webdriver-manager套件

```
!pip install webdriver-manager
```

```
from selenium import webdriver  
from selenium.webdriver.chrome.service import Service  
from webdriver_manager.chrome import ChromeDriverManager  
driver_file = ChromeDriverManager().install()  
driver = webdriver.Chrome(service=Service(driver_file))
```



開始爬蟲

確定網站

■ URL :

https://www.facebook.com/groups/foodspick/?locale=zh_TW&checkpoint_src=any

facebook

鄉民食堂美食社團

加入社團

關於

劉威承
1天 ·

堪稱最強台北長島冰茶酒吧人氣破表啦！遠近馳名的西門町，一直是許多國內外觀光客必訪的指標性景點漫步在各條街道上可見眾多特色餐廳無論是白天或晚上的吃喝玩樂都能一次滿足與朋友們在夜晚的西門小酌絕對不能少！..... 查看更多

DALANG.TW
【冰茶專賣店】台北最強長島冰茶酒吧推薦，世界各地特色融入調酒，大啖懷舊台式古早味美食！ - 達浪的日常手札

1 賽評 留言 分享

Kai Wei Chou
1天 ·

澎湖花火節5/2正式開始～～

近期影音內容

登入或註冊 Facebook 即可和親朋好友以及認識的人聯繫。

登入 或 建立新帳號

利用Python打開網站

```
driver_file = ChromeDriverManager().install()  
driver = webdriver.Chrome(service=Service(driver_file))  
url =  
"https://www.facebook.com/groups/foodspick/?locale=zh_TW&checkpoint_src=an  
y"  
driver.get(url)
```

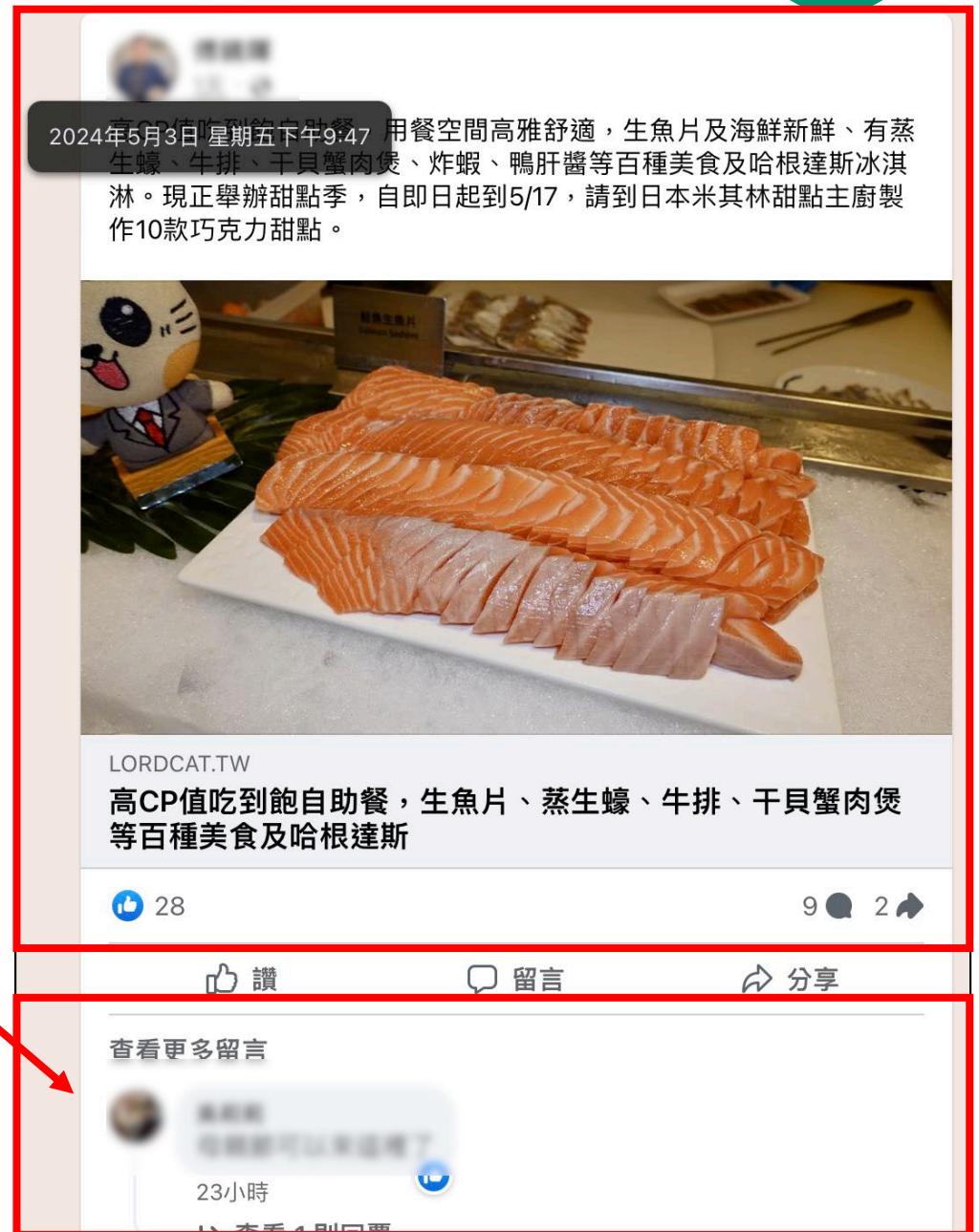
- 社群打開後請手動登陸帳號

爬蟲流程建議

- Step-1: 列出要抓取的資料
- Step-2: 分析Chrome中網頁原始碼的架構
- Step-3: 觀察Chrome中目標資料element的特徵
- Step-4: 搜尋element，搜尋的element的方式非必要一律用CSS Selector
- Step-5: 從element中提取資料

要抓取的資料

- 貼文相關內容
- 回覆相關內容



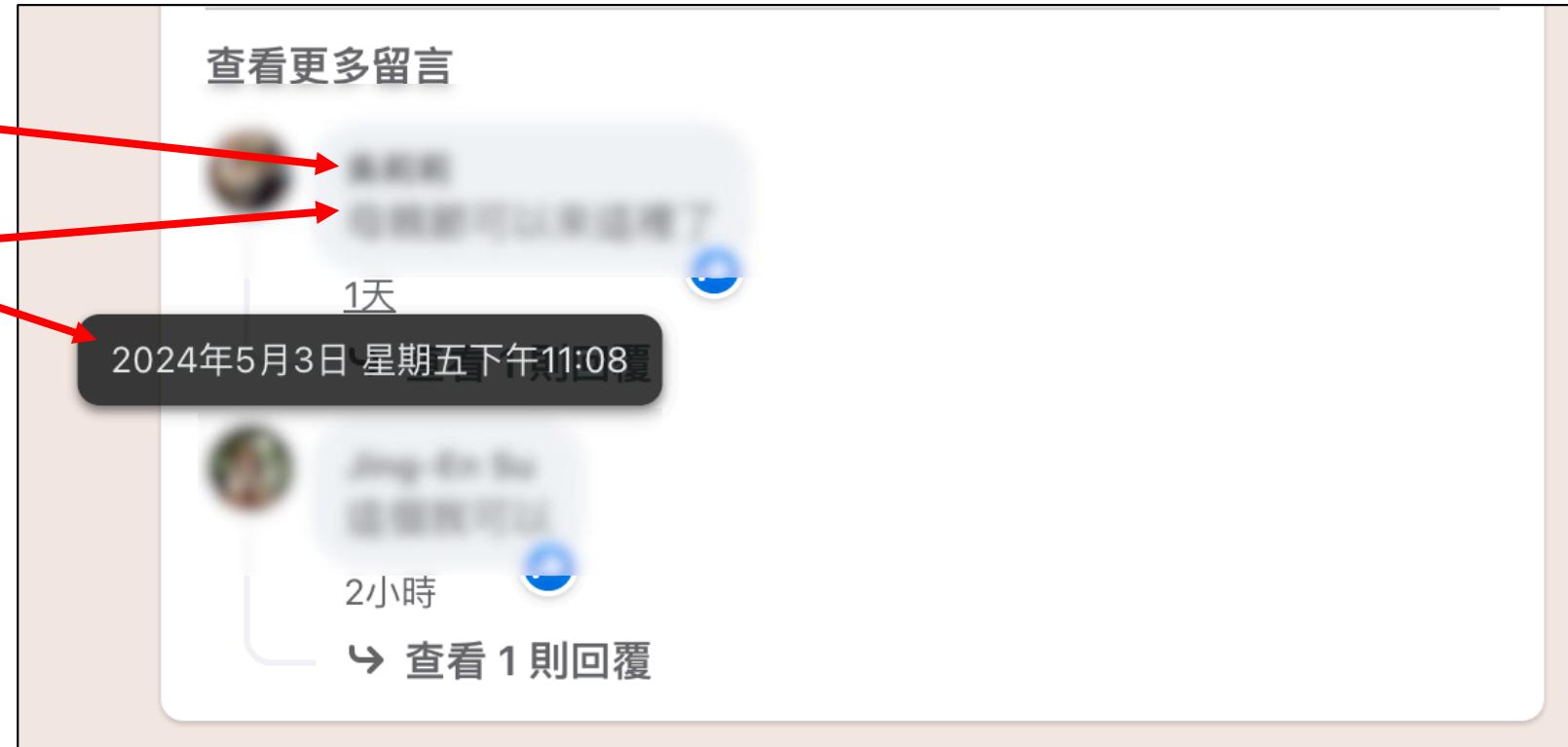
貼文相關內容

- 貼文人
- 貼文日期
- 貼文內容
- 按讚數
- 留言數
- 轉發數



留言相關內容

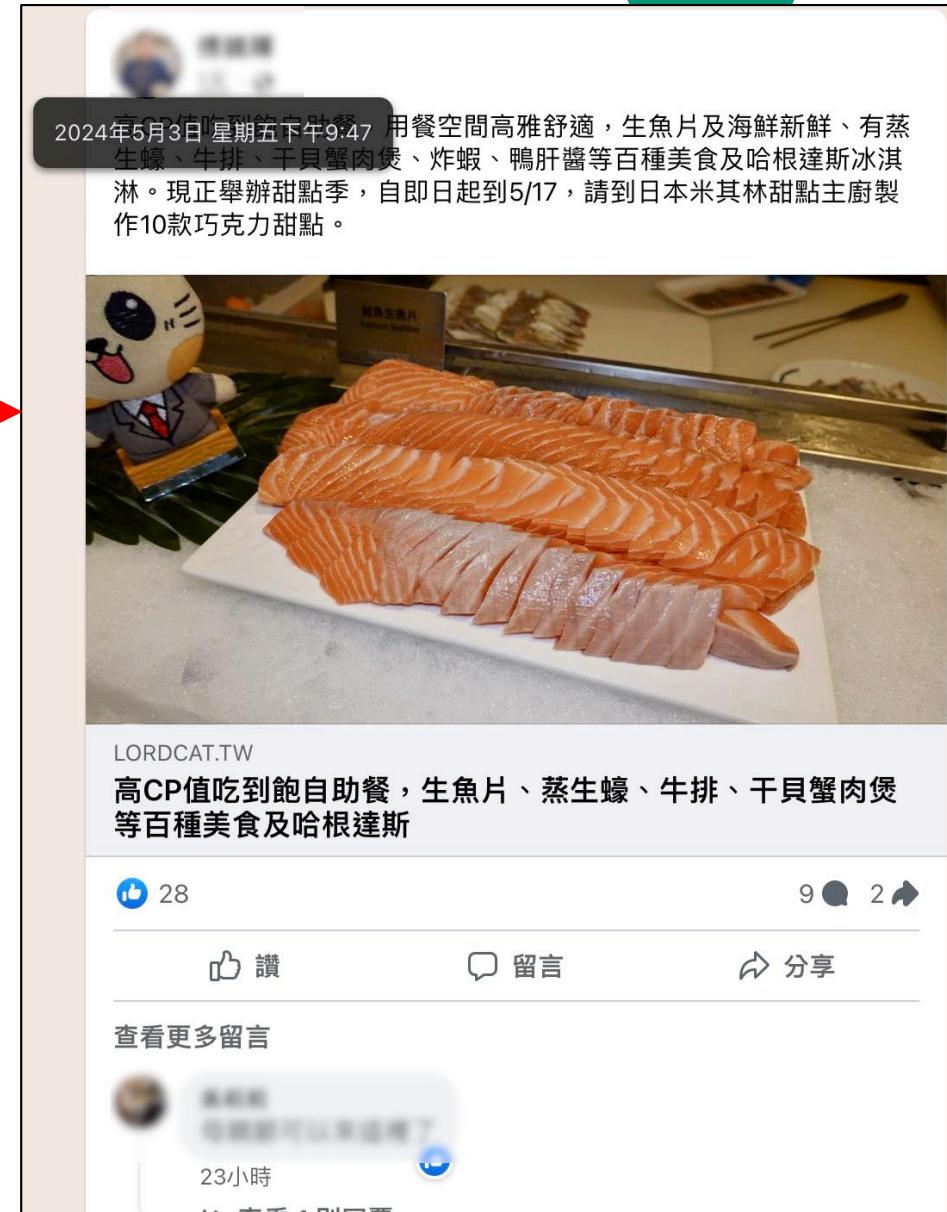
- 留言人
- 留言日期
- 留言內容



Step1: 分析網頁原始碼架構

```
▼<div role="feed">
  <div class="x19f6ikt x169t7cy xsag5q8 x1swvt13"></div>
  ▶<div class="x1yztbdb x1n2onr6 xh8yej3 x1ja2u2z">...</div>
  ▶<div class="x1yztbdb x1n2onr6 xh8yej3 x1ja2u2z">...</div>
```

- 「feed」以下是所有已載貼文
- 從第二行開始是真正貼文



Python Code

```
# 找出已載入的所有貼文(feed)
elements = driver.find_elements(By.CSS_SELECTOR, "div[class='x1yztbdb x1n2onr6 xh8yej3 x1ja2u2z']")
elements
feed = elements[7] # 選出要抓的貼文
move_element_be_visible(driver, feed) # 將貼文移動至可見處
```

Note: move_element_be_visible是自己寫的function

分析網頁原始碼架構

■進一步分析架構，
找到相關內容分別
的位置

```
▼<div class="x1yztbdb x1n2onr6 xh8yej3 x1ja2u2z"> == $0
  ▼<div class="x1n2onr6 x1ja2u2z">
    ▼<div class>
      ▼<div class>
        ▼<div aria-describedby=":r28: :r29: :r2a: :r2c: :r2b:" aria-labelledby=":r27:" class="x1a2a7pz" role="article">
          ▼<div class="x78zum5 xdt5ytf"> flex
            ▼<div class="x9f619 x1n2onr6 x1ja2u2z">
              ▼<div class="x78zum5 x1n2onr6 xh8yej3"> flex
                ▼<div class="x9f619 x1n2onr6 x1ja2u2z x1jx94hy x1qpq9i9 xdney7k xu5ydu1 xt3gfdk xh8yej3 x6ikm8r x10wlt62 xquyuld" style="border-radius: max(0px, min(var(--card-corner-radius), calc((100vw - 4px - 100%) * 9999))) / var(--card-corner-radius);">
                  ▼<div>
                    <div></div>
                  ▼<div>
                    ▼<div class="html-div xe8uvvx xdj266r x11i5rnm xat24cr x1mh8g0r xexx8yu x4uap5 x18d9i69 xkh6sd">
                      ▼<div class="html-div xe8uvvx xdj266r x11i5rnm xat24cr x1mh8g0r xexx8yu x4uap5 x18d9i69 xkh6sd">
                        <div class></div>
                        ▶<div class="html-div xe8uvvx xdj266r x11i5rnm xat24cr x1mh8g0r xexx8yu x4uap5 x18d9i69 xkh6sd">...</div>
                        ▶<div class="html-div xe8uvvx xdj266r x11i5rnm xat24cr x1mh8g0r xexx8yu x4uap5 x18d9i69 xkh6sd">...</div>
                        ▶<div>...</div>
```

分析網頁原始碼架構

```
►<div class="html-div xe8uvvx xdj266r x11i5rnm xat24cr x1mh8g0r xexx8yu x4uap5 x18d9i69 xkh6sd">...</div>
►<div class="html-div xe8uvvx xdj266r x11i5rnm xat24cr x1mh8g0r xexx8yu x4uap5 x18d9i69 xkh6sd">...</div>
►<div>...</div>
```

高CP值吃到飽自助餐，用食空間高雅舒適，生魚片及海鮮新鮮、有蒸生蠔、牛排、干貝蟹肉煲、炸蝦、鴨肝醬等百種美食及哈根達斯冰淇淋。現正舉辦甜點季，自即日起到5/17，請到日本米其林甜點主廚製作10款巧克力甜點。

LORDCAT.TW
高CP值吃到飽自助餐，生魚片、蒸生蠔、牛排、干貝蟹肉煲等百種美食及哈根達斯

28 9 2

讚 留言 分享

查看更多留言

貼文人



```
▼<h2 id=":r4b:" class="html-h2 xe8uv  
vx x11i5rnm xat24cr x1mh8g0r xexx8y  
u x4uap5 x18d9i69 xkhd6sd x1vvkbs x  
1heor9g x1qlqyl8 x1pd3egz x1a2a7pz  
x1gslohp x1yc453h">  
▼<span class="xt0psk2">
```

```
# 貼文發佈人  
feed_man = feed.find_elements(By.CSS_SELECTOR,  
"h2>span[class='xt0psk2'])[0].text  
print(feed_man)
```

貼文時間



- 滑鼠要移動到上面才會顯示
- 顯示的element位置不確定

```
# 貼文時間  
locator = "div[class='xu06os2 x1ok221b'] > span[dir='auto'] > span > span> span > a[role='link'][tabindex='0']"  
element = feed.find_elements(By.CSS_SELECTOR, locator)[0]  
mouse_move_to(driver, element)  
time.sleep(1) # 等待網頁反應  
feed_time = driver.find_elements(By.CSS_SELECTOR, "div[role='tooltip']")[0].text  
print(feed_time)
```

貼文內容

高CP值吃到飽自助餐，用餐空間高雅舒適，生魚片及海鮮新鮮、有蒸生蠔、牛排、干貝蟹肉煲、炸蝦、鴨肝醬等百種美食及哈根達斯冰淇淋。現正舉辦甜點季，自即日起到5/17，請到日本米其林甜點主廚製作10款巧克力甜點。

貼文內容

```
locator = "div[dir='auto'] > div[class='x1iorvi4 x1pi30zi x1l90r2v x1swvt13'] > span[dir='auto']"  
  
feed_content = feed.find_elements(By.CSS_SELECTOR,  
locator)[0].text  
  
print(feed_content)
```



A screenshot of a browser's developer tools showing the HTML structure of the post content. A red arrow points from the text area in the code block above to the corresponding span element in the DOM tree.

```
<div dir="auto" class="html-div xe8uv vx xdj266r x11i5rnm xat24cr x1mh8g0r xexx8yu x4uap5 x18d9i69 xkhd6sd">  
  <div class="x1iorvi4 x1pi30zi x1l90r2v x1swvt13" id=":r84:"> == $0  
    <span class="x193iq5w xeuugli x13faqbe x1vvkbs xlh3980 xvmahel x1n0sxbx x1lliihq x1s928wv xhkezso x1gmr53x x1cpjm7i x1fgarty x1943h6 x4zkp8e x3x7a5m x6prxxf xvq8zen x01l8bm xzsf02u" dir="auto">  
      ::before  
    <div class="x1e56ztr">  
      <span class="x193iq5w xeuugli x13faqbe x1vvkbs xlh3980 xvmahel x1n0sxbx x6prxxf xvq8zen x01l8bm xzsf02u">  
        "高CP值吃到飽自助餐，用餐空間高雅舒適，生魚片及海鮮新鮮、有蒸生蠔、牛排、干貝蟹肉煲、炸蝦、鴨肝醬等百種美食及哈根達斯冰淇淋。現正舉辦甜點季，自即日起到5/17，請到日本米其林甜點主廚製作10款巧克力甜點。"  
    </span>
```

連結內容

```
# 連結
```

```
element =  
feed.find_elements(By.CSS_SELECTOR,  
"div[class='xmjcpbm x1n2onr6'] > div >  
a[rel='nofollow  
noreferrer'][role='link'][tabindex='0']]")[0]  
  
feed_link = element.get_attribute('href')  
  
feed_link
```



LORDCAT.TW

高CP值吃到飽自助餐，生魚片、蒸生蠔、牛排、干貝蟹肉煲等百種美食及哈根達斯

```
# 連結文字
```

```
element = feed.find_elements(By.CSS_SELECTOR,  
"span[dir='auto'] > span > span[dir='auto'])")[0]  
  
feed_link_alt = element.text  
  
feed_link_alt
```

連結內容

```
# 照片  
feed_pic = feed.find_elements(By.CSS_SELECTOR,  
"div[class='x10l6tqk x13vifvy'] >  
img")[0].get_attribute('src')  
  
feed_pic
```



LORDCAT.TW

高CP值吃到飽自助餐，生魚片、蒸生蠔、牛排、干貝蟹肉煲等百種美食及哈根達斯

按讚數



33

9

2



```
▼<div class="x6s0dn4 x78zum5 x1iyj
qo2 x6ikm8r x10wlt62"> flex
►<span aria-label="看看誰對這個傳達了心情" class="x1ja2u2z" role="toolbar">...</span>
<div class="x9f619 x1ja2u2z xzp
qnlu x1hyvwdk xjm9jq1 x6ikm8r x
10wlt62 x10l6tqk x1i1rx1s">所有
心情 : </div>
▼<span aria-hidden="true" class=
"rbpxyo x6ikm8r x10wlt62 xlyip
yv x1exxlbk">
▼<span>
  <span class="x1e558r4">33
  </span> == $0
</span>
</span>
</div>
```

按讚數

```
element = feed.find_elements(By.CSS_SELECTOR,
"span[class='x1e558r4'])[0]
```

```
emotion = element.text
```

```
emotion
```

留言數

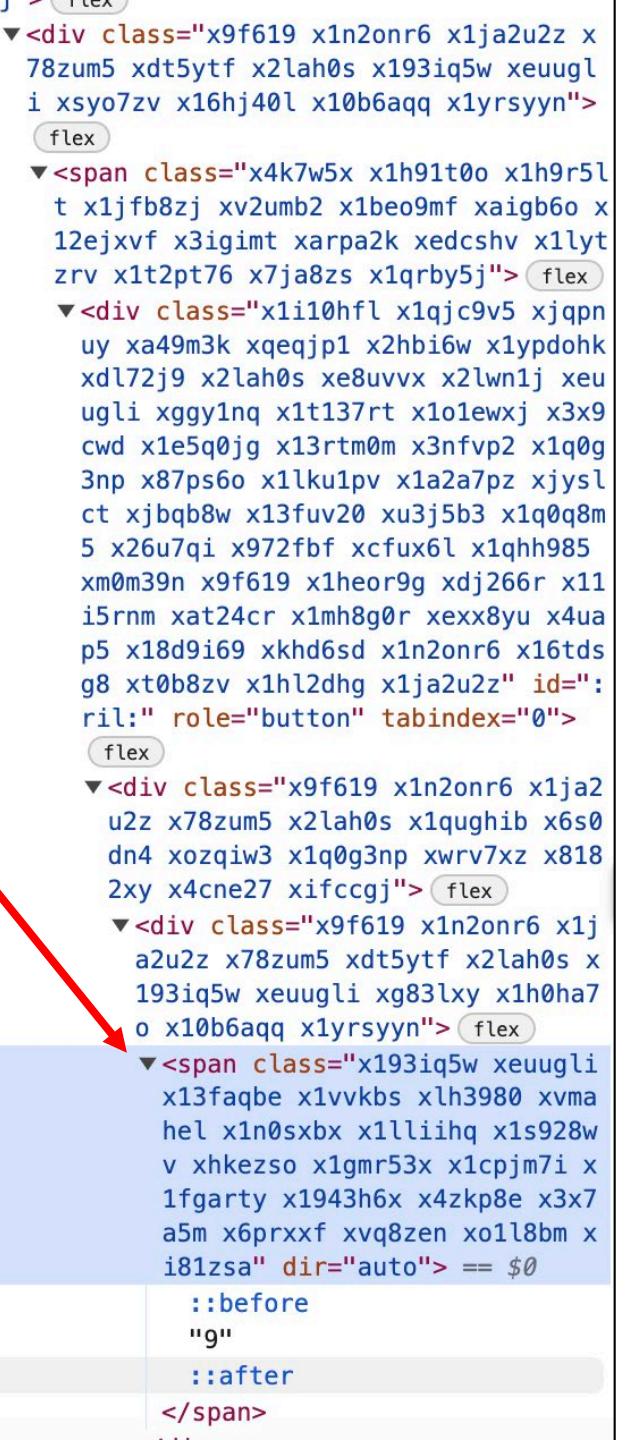


33

9

2

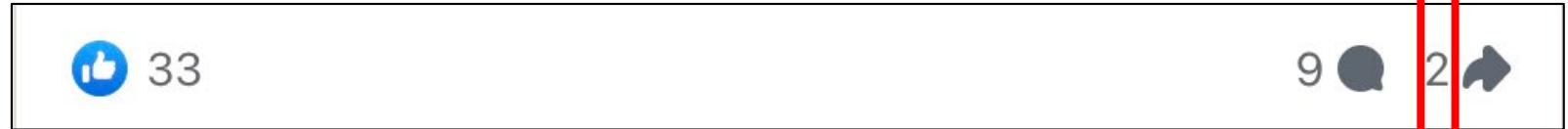
```
# 留言數  
element = feed.find_elements(By.CSS_SELECTOR, "div  
> span > div[role='button'][tabindex='0'] > div > div  
> span[dir='auto'])[0]  
  
comment_number = element.text  
  
comment_number
```



The screenshot shows the browser's developer tools inspecting an element. A red box highlights the number '9' in a white rectangular box, which is part of a larger element structure. A red arrow points from the highlighted area in the screenshot to the corresponding DOM node in the developer tools. The DOM node is a `span` element with class `x193iq5w xeuugli x13faqbe x1vvkbs xlh3980 xvma hel x1n0sxbx x1lliihq x1s928w v xhkezso x1gmr53x x1cpjm7i x1fgarty x1943h6x x4zkp8e x3x7a5m x6prxxf xvq8zen x01l8bm x181zsa` and attribute `dir="auto"`. The text '9' is shown under the `::before` pseudo-element.

```
j > flex  
▼<div class="x9f619 x1n2onr6 x1ja2u2z x78zum5 xdt5ytf x2lah0s x193iq5w xeuugli xsys07zv x16hj40l x10b6aqq x1yrsyn">  
  flex  
  ▼<span class="x4k7w5x x1h91t0o x1h9r5lt x1jfb8zj xv2umb2 x1beo9mf xaigb6o x12ejxvf x3igimt xarpa2k xedcshv x1lytzrv x1t2pt76 x7ja8zs x1qrby5j"> flex  
  ▼<div class="x1i10hfl x1qjc9v5 xjqpnu xxa49m3k xqeojp1 x2hbi6w x1ypdohk xdl72j9 x2lah0s xe8uvvx x2lwn1j xeuugli xggyl1nq x1t137rt x1o1ewxj x3x9 cwd x1e5q0jg x13rtm0m x3nfvp2 x1q0g3np x87ps6o x1lku1pv x1a2a7pz xjyslct xjbqb8w x13fuv20 xu3j5b3 x1q0q8m5 x26u7qi x972fbf xcfux6l x1qhh985 xm0m39n x9f619 x1heor9g xdj266r x11i5rnm xat24cr x1mh8g0r xexx8yu x4uap5 x18d9i69 xkh6sd x1n2onr6 x16tdsg8 xt0b8zv x1hl2dhg x1ja2u2z" id=":ril:" role="button" tabindex="0">  
    flex  
  ▼<div class="x9f619 x1n2onr6 x1ja2u2z x78zum5 x2lah0s x1qughib x6s0dn4 xozqiw3 x1q0g3np xwrv7xz x8182xy x4cne27 xifccgj"> flex  
  ▼<div class="x9f619 x1n2onr6 x1ja2u2z x78zum5 xdt5ytf x2lah0s x193iq5w xeuugli xg83lxy x1h0ha7o x10b6aqq x1yrsyn"> flex  
  ▼<span class="x193iq5w xeuugli x13faqbe x1vvkbs xlh3980 xvma hel x1n0sxbx x1lliihq x1s928w v xhkezso x1gmr53x x1cpjm7i x1fgarty x1943h6x x4zkp8e x3x7a5m x6prxxf xvq8zen x01l8bm x181zsa" dir="auto"> == $0  
    ::before  
    "9"  
    ::after  
  </span>
```

轉發數



```
# 轉發數
```

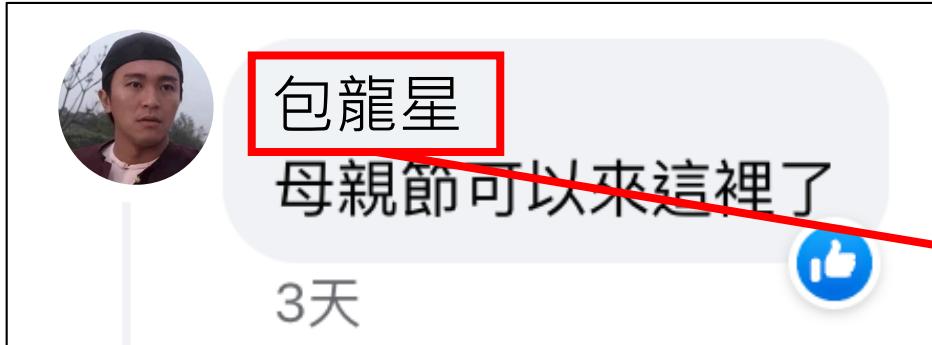
```
element = feed.find_elements(By.CSS_SELECTOR, "div > span > div[role='button'][tabindex='0'] > div > div > span[dir='auto'])[1]
```

```
comment_number = element.text
```

```
comment_number
```

```
<div class="x9f619 x1n2onr6 x1ja2u2z x78zum5 xdt5ytf x2lah0s x193iq5w xeuugli xsys07zv x16hj40l x10b6aqq x1yrsyyn">
  flex
    <span class="x4k7w5x x1h91t0o x1h9r51t x1jfb8zj xv2umb2 x1beo9mf xaigb6o x12ejxvf x3igimt xarpa2k xedcshv x1lytzrv x1t2pt76 x7ja8zs x1qrby5j"> flex
      <div class="x1i10hfl x1qjc9v5 xjqpnuy xa49m3k xqeojp1 x2hbi6w x1ypdohk xdl72j9 x2lah0s xe8uvvx x2lwn1j xeuugli xggy1nq x1t137rt x1o1ewxj x3x9 cwd x1e5q0jg x13rtm0m x3nfvp2 x1q0g3np x87ps6o x1lku1pv x1a2a7pz xjyslct xjbqb8w x13fuv20 xu3j5b3 x1q0q8m5 x26u7qi x972fbf xcfux6l x1qhh985 xm0m39n x9f619 x1heor9g xdj266r x11i5rnm xat24cr x1mh8g0r xexx8yu x4ua p5 x18d9i69 xkhd6sd x1n2onr6 x16tdsg8 xt0b8zv x1hl2dhg x1ja2u2z" role="button" tabindex="0"> flex
        <div class="x9f619 x1n2onr6 x1ja2u2z x78zum5 x2lah0s x1qughib x6s0dn4 xozqiw3 x1q0g3np xwrv7xz x8182xy x4cne27 xifccgj"> flex
          <div class="x9f619 x1n2onr6 x1ja2u2z x78zum5 xdt5ytf x2lah0s x193iq5w xeuugli xg83lxy x1h0ha7o x10b6aqq x1yrsyyn"> flex
            <span class="x193iq5w xeuugli x13faqbe x1vvkbs xlh3980 xvma hel x1n0sxbx x1lliihq x1s928wv xhkezso x1gmrs53x x1cpjm7i x1fgarty x1943h6x x4zkp8e x3x7a5m x6prxxf xvq8zen xo1l8bm xi81zsa" dir="auto"> == $0
              ::before
              "2"
              ::after
            </span>
```

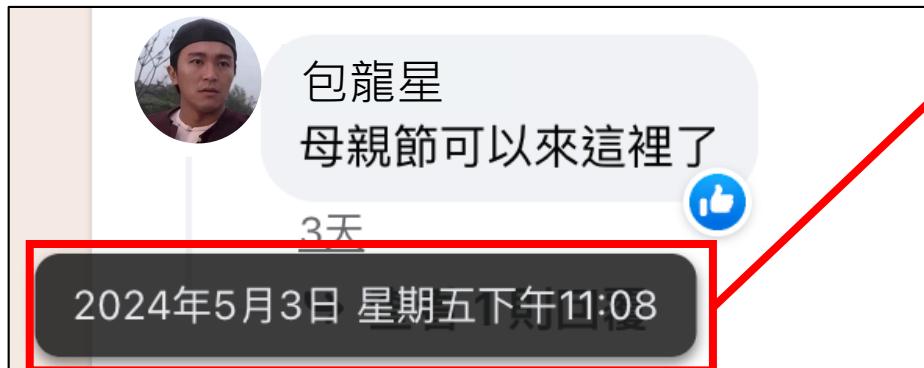
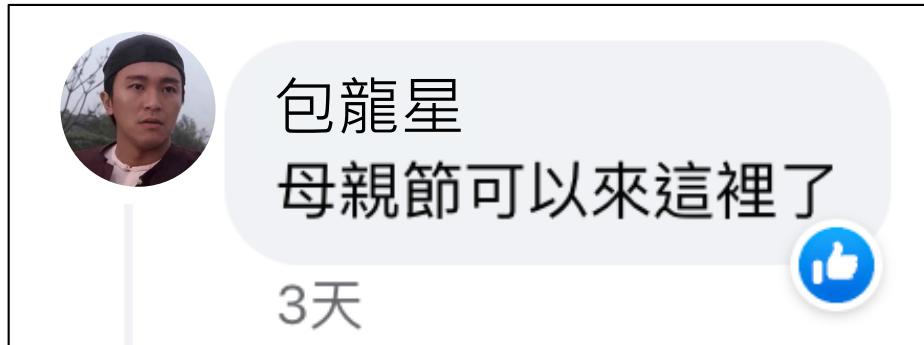
留言人



```
# 留言人
element = feed.find_elements(By.CSS_SELECTOR,
"span[class='x3nfvp2'])[0]
comment_man = element.text
comment_man
```

```
▼<span class="xt0psk2">
  ▼<div aria-hidden="false" class="x
    1i10hfl xjbqb8w x1ejq31n xd10rxx
    x1sy0etr x17r0tee x972fbf xcfux6l
    x1qhh985 xm0m39n x9f619 x1ypdohk
    xt0psk2 xe8uvvx xdj266r x11i5rnm
    xat24cr x1mh8g0r xexx8yu x4uap5 x
    18d9i69 xkh6sd x16tdsg8 xggy1nq
    x1a2a7pz x1heor9g xt0b8zv x1hl2dh
    g" role="button" tabindex="0">
    ▼<span class="x3nfvp2" flex>
      ▼<span class="x193iq5w xeuugli
        x13faqbe x1vvkbs xlh3980 xvma
        hel x1n0sxbx x1lliihq x1s928w
        v xhkezso x1gmr53x x1cpjm7i x
        1fgarty x1943h6x x4zkp8e x676
        frb x1nxh6w3 x1sibtaa x1s688f
        xzsf02u" dir="auto"> == $0
          ::before
          "吳莉莉"
          ::after
        </span>
```

留言時間

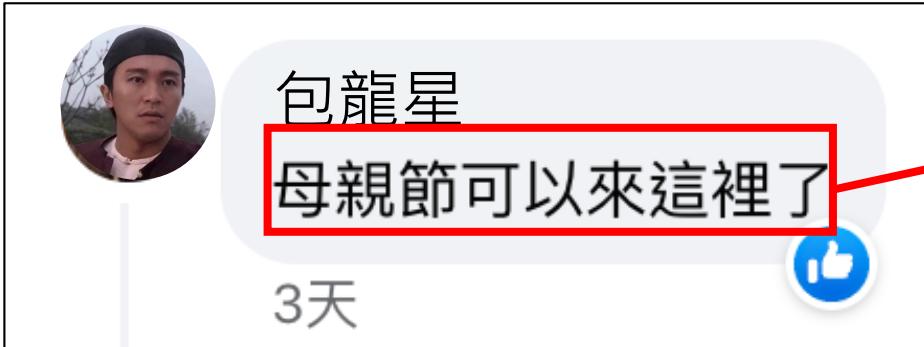


```
<div class="x6s0dn4 x3nfvp2" flex>
  <ul aria-hidden="false" class="x1n0m28w x78zum
    5 x1wfe3co xat24cr xsgj6o6 x1o1nzlu xyqdw3p">
    flex
    <li class="x1rg5ohu x1emribx x1i64zmx">
      <span class="x4k7w5x x1h91t0o x1h9r5lt x1j
        fb8zj xv2umb2 x1beo9mf xaigb6o x12ejxvf x3
        igimt xarpa2k xedcshv x1lytzrv x1t2pt76 x7
        ja8zs x1qrby5j">
        <div>
          <a class="x1i10hfl xjbqb8w x1ejq31n xd
            10rxx x1sy0etr x17r0tee x972fbf xcfux6
            l x1qhh985 xm0m39n x9f619 x1ypdohk xt0
            psk2 xe8uvvx xdj266r x11i5rnm xat24cr
            x1mh8g0r xexx8yu x4uap5 x18d9i69 xkh6
            sd x16tdsg8 xggyl1nq x1a2a7pz xt0b8zv x
            1hl2dhg xi81zsa xo1l8bm" href="https://
            www.facebook.com/groups/foodspick/pos
            ts/2112247919140154/?c...mpRPEExHbZQ7PuVC
            KVI_bye8uaFMfkclVpYvYfiPsjxZKbcW6rMIs0
            90e0pwWx& tn =R]-R" role="link"
            tabindex="0">
            <span class="html-span xdj266r x11i5r
            nm xat24cr x1mh8g0r xexx8yu x4uap5 x1
            8d9i69 xkh6sd x1hl2dhg x16tdsg8 x1vv
            kbs">3天</span> == $0
```

Python Code

```
# 留言時間  
element = feed.find_elements(By.CSS_SELECTOR, "div[class='x6s0dn4  
x3nfvp2'][0]  
mouse_move_to(driver, element)  
time.sleep(1) # 等待網頁反應  
comment_time = driver.find_elements(By.CSS_SELECTOR,  
"div[role='tooltip'][0].text  
print(comment_time)
```

留言內容



```
# 留言內容
element =
feed.find_elements(By.CSS_SELECTOR,
"div[class='x1lliihq xjkvuk6 x1iorvi4'][0]")
comment_content = element.text
comment_content
```

```
<div class="x1lliihq xjkvuk6 x1iorvi4"> == $0
  <span class="x193iq5w xeuugli x13fa qbe x1vvkbs xlh3980 xvmahel x1n0sxb x x1lliihq x1s928wv xhkezso x1gmr53 x x1cpjm7i x1fgarty x1943h6x x4zkp8 e x3x7a5m x6prxxf xvq8zen xo1l8bm x zsf02u" dir="auto" lang="zh-TW">
    ::before
    <div class="xdj266r x11i5rnm xat24cr x1mh8g0r x1vvkbs">
      <div dir="auto" style="text-align: start;">母親節可以來這裡了
      </div>
```

資訊整理與輸出

```
# 資訊整理並輸出CSV  
df_feeds = pd.DataFrame({"Feed Number": [1],  
    "Feed Man": [feed_man],  
    "Feed Time": [feed_time],  
    "Feed Content": [feed_content],  
    "Feed Content Picture": [feed_pic],  
    "Feed Link String": [feed_link_alt],  
    "Feed Link": [feed_link],  
    "Comment Man": [comment_man],  
    "Comment Time": [comment_time],  
    "Comment Content": [comment_content]})  
df_feeds.to_csv("../feeds.csv", encoding="utf-8-sig", index=False)
```



The End

Thank you for your attention!

Email: chchan@ntnu.edu.tw

Web: toodou.github.io

